

Gradual or abrupt? The phonetic path to morphologisation

Patrycja Strycharczuk and James M. Scobbie

To appear in *Journal of Phonetics*

Abstract

While some sound changes occur in environments defined in purely phonological terms, others may become sensitive to morphological boundaries. In this paper, we investigate the phonetic nature of this latter diachronic development: does it happen through small gradient increments, or is there a categorical shift from one allophone to another? We focus on GOOSE-fronting and /l/-darkening in Southern British English, the interaction of which is sensitive to morphological boundaries. Relatively retracted realisations of the vowel and dark realisations of the /l/ appear before a morpheme boundary, even when a vowel follows (e.g. *fool-ing*), whereas in monomorphemic words (e.g. *hula*), there is more /u:/-fronting, and the /l/ is relatively lighter. We analyse the phonetic realisation of such pairs as *hula* vs. *fool-ing* in 20 speakers of Southern British English using both acoustic and articulatory (ultrasound) instrumental methods. All the speakers express the morphological contrast in some way, although effect sizes vary dramatically. For some speakers, the contrast involves subtle articulatory differences without any clear acoustic consequences, whereas other speakers show robust differences employing multiple phonetic correlates. We therefore argue that the *hula*~*fool-ing* contrast would be misrepresented if analysed in terms of a small number of either /u:/ or /l/ allophonic categories. Instead, we interpret the results as supporting the predictions of phonological frameworks that incorporate phonetically-gradient morphologisation.

1 Introduction

At some point of their development, sound changes may start showing morphological effects. For example, in most modern-day English dialects, /g/ has been lost from /ŋg/ clusters at the end of the word (*sing*), as well as preceding a morpheme boundary (*sing-er*), yet /g/ is still pronounced before a vowel when no morpheme boundary intervenes (*finger*). Similar cases of morphological conditioning abound in English dialects, affecting for instance, /l/-darkening in American English (Boersma & Hayes, 2001; Lee-Kim et al., 2013) and the Scottish Vowel Length Rule (Aitken, 1981; Scobbie et al., 1999; Scobbie & Stuart-Smith, 2008). What has been a topic of debate is whether morphological effects are pervasive in language, or whether they only emerge at a defined point once a sound change is firmly established.

So far, empirical observations concerning morphological constraints on phonological processes have only been made for relatively advanced sound changes. Some authors explicitly argue that this is not coincidental, and that morphologisation is a late development in sound change (Bermúdez-Otero & Trousdale, 2012; Ramsammy, 2015).¹ This argument grows out of a modular view of grammatical architecture, in which sound change is analysed as a progression through grammatical hierarchy, involving some abrupt categorical shifts. In this architecture, morphological information is not accessible to phonetics, and therefore the prediction follows that a sound change in its earliest phases cannot be sensitive to morphological boundaries, but must be purely phonetic, sensitive to at most a phonological environment. A gradient phonetic process may, however, undergo stabilisation, where it is re-interpreted as a phrase-level categorical rule that computes phonologically distinct allophones. It is then and only then that further change might occur, restricting the domain of the rule application to progressively more narrow domains. For instance, a rule may progress from being phrase-final to being word-final (this change is called domain narrowing I in Bermúdez-Otero & Trousdale’s

¹We use the term ‘morphologisation’ referring to the emergence of morphological effects in sound change, such as the difference between monomorphemic *finger* and morphologically complex *sing-er*. Note that this is different from the terminology used by Bermúdez-Otero & Trousdale (2012) and Ramsammy (2015), who call this change ‘domain narrowing II’.

life-cycle model), and subsequently, also to being morpheme-final (domain narrowing II). The second stage of domain narrowing may result in fuzzy contrasts, such as *finger*~*sing-er*, but crucially, the final /g/ in *sing* must have been subject to categorical deletion before the *finger*~*sing-er* contrast appears, because phonetics-morphology interactions are necessarily mediated via intermediate phonological representation. Note that this prediction is not unique to the model developed by Bermúdez-Otero & Trousdale (2012) and Ramsammy (2015). Rather, the model is consistent with more general predictions concerning the absence of morphology-phonetics interactions that follow from modular approaches to grammatical architecture (e.g. Levelt et al. 1999; see also Kiparsky 1985 for discussion on modularity effects and distinction between lexical and post-lexical phonology).

In contrast, proponents of an alternative, non-modular, view of grammatical architecture have challenged the prediction concerning the absence of direct interaction between morphology and phonetics. Bybee (2001) argues for a model where the phonetic realisation of words can be encoded as a cloud of variants associated directly with each lexical item, which may, moreover, be more or less explicitly compositional in the way it encodes morphological relationships, rather than being computed from a phonological representation derived in its turn by a system of morphological combinatorics. Apparent structural effects follow from connections between related words. For instance, the diachronic deletion of /g/ in *singer* would be attributed to analogy or shared activation (*sing-er* being in the same paradigm as *sing*), rather than to the presence of a morphological boundary. An important aspect of this proposal involves the storage of phonetic detail for individual words. The combination of such concept of storage and the nature of lexical analogy yields a prediction that apparent morphological effects may emerge at any point in sound change. (Bybee, 2001, p.68) makes this explicit, stating that “(...) morphologization occurs very gradually and much earlier than generally supposed”. Bybee’s comment on early morphologisation alludes to the possibility that morphological effects may initially be too small to be immediately observable, in either absolute or relative terms.

We have already mentioned that morphological effects have so far only been reported for relatively advanced sound changes. However, it is not entirely clear whether this is really a matter of morphological effects first appearing, or morphological effects becoming big enough for native speakers, or linguists, to notice. The majority of the documented cases of morphologisation rely on transcription data, which are limited by the categorical nature of segmental broad transcription, by categorical speech perception and meta-phonological awareness. Therefore, we must also consider the possibility that morphological structure may interact with small-scale phonetic processes, but that those interactions are not very prominent, and as such they are missed by linguistic descriptions.

A considerable body of research has developed in this particular area, but the results so far remain inconsistent. There is some evidence that the presence of a morphological boundary may increase acoustic duration of the preceding rhyme, although Sugahara & Turk (2009) only observe this effect in slow speech. Similarly, the duration of the plural suffix in English has been reported to differ from the duration of word-final /s,z/ in monomorphemic words. Note, however, that while Song et al. (2013a) report increased duration for the suffix, Plag et al. (2015) find the opposite effect. For Korean, Cho & Keating (2001) report increased variability in gestural timing in morphologically complex words compared to monomorphemes. Beňuš (2012) presents evidence that yer vowels in Slovak, i.e. vowels that alternate with zero within the morphological paradigm, are phonetically weaker than non-yer vowels that are not subject to similar alternations. Song et al. (2013b) find evidence of tongue lowering during /k/ in monomorphemic *box*, compared to bimorphemic *rocks*, a difference they ascribe to increased anticipatory co-articulation in tautomorphemic clusters. However, it has to be noted that pairs such as *box* and *rocks* also differ in their lexical frequency (*rocks* is overall more frequent), in orthography (*rocks* is spelled with more letters), and in their onset consonant. All of these factors may exert subtle effects on articulation. In an articulatory study that controlled for frequency, spelling and segmental effects through use of nonce words, Mousikou et al. (2015) do not find evidence for increased coartic-

ulation in tautomorphemic clusters compared to heteromorphemic ones, although the authors express reservations about the validity of their null result, given some limitations in statistical power.

Potential confounds make it challenging to study the effect of morphology on pronunciation, especially since the putative morphological effects may be very small. Sound change in progress is an important source of evidence in this context, since effect sizes are by definition on an increasing trajectory in sound change. In addition, competing theories of grammatical architecture make diverging predictions concerning the phonetic nature of ongoing morphologisation. We shall illustrate this, using an example from GOOSE-fronting before /l/ which leads to a difference in the degree of fronting in words such as monomorphemic *hula* and morphologically complex *fool-ing*.² The GOOSE vowel has been undergoing fronting in numerous dialects of English, including the varieties treated as standard in England (Bauer, 1985; Hawkins & Midgley, 2005; McDougall & Nolan, 2007; Harrington, 2007; Harrington et al., 2008; Chládková & Hamann, 2011). However, this fronting process has apparently not occurred in words where the GOOSE vowel is followed by a coda /l/, such as *fool*. Furthermore, similar blocking of GOOSE-fronting has also been noted preceding morpheme-final /l/, e.g. in *fool-ing* (see below). In contrast, GOOSE-fronting has affected words where the following /l/ is not morpheme-final, e.g. in *hula*.

According to non-modular views of grammatical architecture, the morphologically-conditioned difference in the degree of GOOSE-fronting between *hula* and *fool-ing* may be present as soon as any degree of GOOSE-fronting occurs in *hula*. This is because vowel fronting in *fool-ing* will be partially limited through analogy to *fool*, where the fronting does not apply, due to presence of a dark /l/ in the syllable rhyme. In terms of segmental and prosodic context however, *fool-ing* is more similar to *hula* than to *fool*, as the /u:l/ sequence occurs word medially before a vowel. These opposing influences

²Although we use the specific example of GOOSE-fronting, the predictions hold for the development of any phonetic category, and they also extend to /l/-darkening, which we shall discuss in more detail in Sections 3 and 4. Note that the interaction between /l/-darkening and GOOSE-fronting makes the *hula*~*fool-ing* case more complex than cases like *finger*~*sing-er* or other morphologically-conditioned contrasts we are familiar with.

may create intermediate fronting in morphologically complex words like *fool-ing*. As the GOOSE-fronting in *hula* progresses over time, the morphological distinction between *hula* and *fool-ing* becomes phonetically more extreme, giving rise to perceivable differences in fronting between monomorphemic *hula* and bimorphemic *fool-ing*. This gradual phonetic development of morphologisation is schematised in Figure 1.

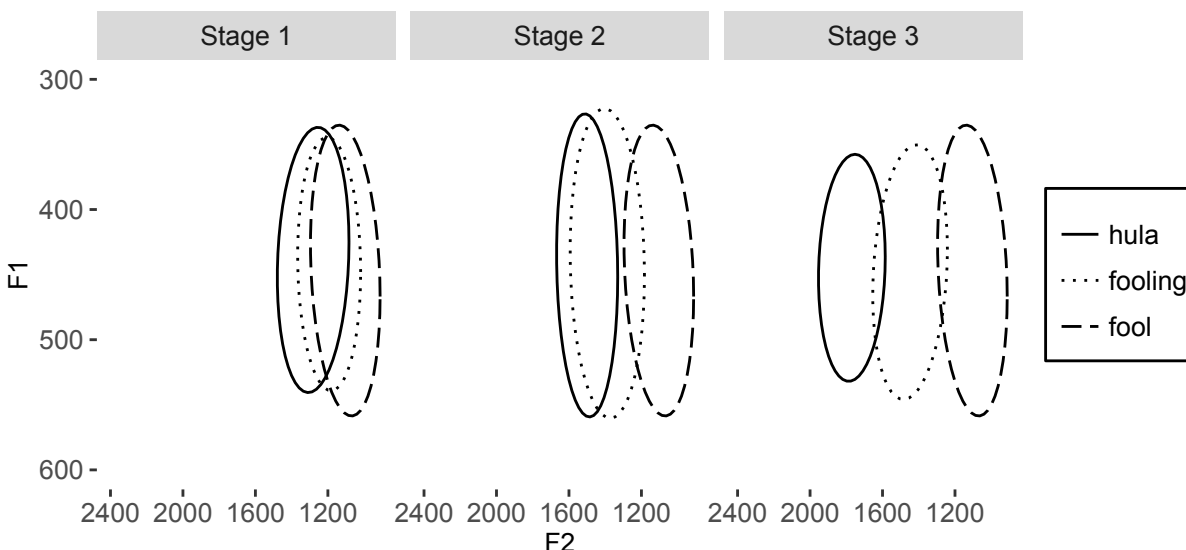


Figure 1: Gradient development of the increasingly different degrees of GOOSE-fronting in *hula* and *fool-ing*.

Figure 2 illustrates an alternative set of predictions which follow from modular views of grammatical architecture. In this architecture, morphological information may affect phonological operations, but it is not directly accessible to phonetics. Therefore, no morphological effect is expected in early phonetic stages of sound change; initially, GOOSE-fronting should affect *hula* and *fool-ing* equally. A morphological effect may emerge only after the blocking of GOOSE-fronting has become an allophonic phonological rule. A rule like this would generate a front allophone in *hula*, but a back one in *fool*. Once two distinct allophones have emerged, morphologisation may follow. At this point, the rule that blocks GOOSE-fronting before a coda /l/ will also affect /u:l/ sequences at the end of a morpheme, e.g. in *fool-ing*. This re-analysis would introduce an abrupt phonetic shift to a back allophone in *fool-ing*. Note that this does not necessarily entail

that the /u:/ in *fool-ing* and *fool* become identical – the prosodic difference between the two forms (i.e. word boundary following /l/ in *fool*) may condition increased /l/-darkening, and more retracted quality of the vowel in *fool* compared to *fool-ing*. Thus, the allophonic shift in *fool-ing* should be understood as *fool-ing* becoming similar (but not identical) to *fool* in terms of the quality of the /u/ and the /l/. The crucial prediction of this model concerns the distinction between *hula* and *fool-ing*: there is either no distinction whatsoever, or a robust, categorical one.

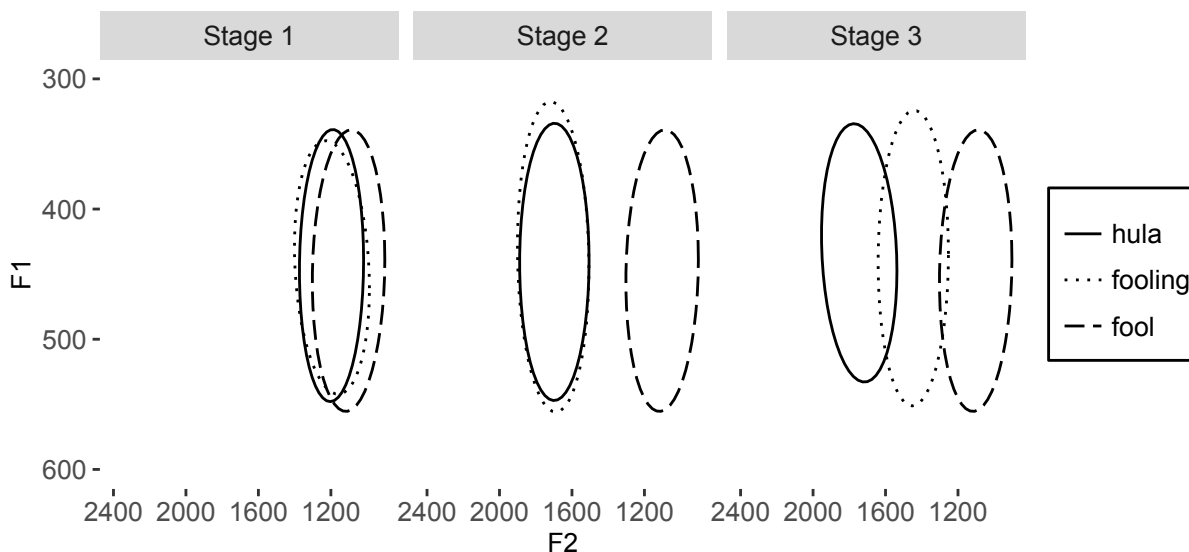


Figure 2: Abrupt development of the GOOSE-fronting contrast between *hula* and *fool-ing*.

In this paper, we investigate the phonetics of developing morphological contrasts in the context of the two diachronic scenarios sketched out above. We present data on the realisation of monomorphemic words like *hula*, and morphologically-complex words like *fool-ing* by speakers of Southern British English (SBE). We focus on this variety because of recent reports that GOOSE-fronting applies to a different degree before /l/, depending on the morphosyntactic context, leading to the emergence of near-minimal pairs such as the opposition between *ruler*, meaning ‘measuring device’, and *rul-er*, meaning ‘political leader’ (Uffmann, 2012). We use an apparent-time design, with a view to finding instances of the *hula*~*fool-ing* contrast at different stages of its development. We expect variation, since GOOSE-fronting is still ongoing in this dialect (see Harrington et al. 2008

and Chládková et al. (2016) for recent findings showing apparent-time differences in the degree of GOOSE-fronting in SBE). As the speakers in our study represent two age groups that are not very far apart (under 28 and over 45), we do not necessarily expect to find apparent-time effects that exactly line up with specific diachronic stages schematised in Figure 1 (for the non-modular models) or Figure 2 (for the modular models). Instead, we ask a question that concerns effect sizes associated with the *hula*~*fool-ing* contrast within speaker: for any speaker who has the *hula*~*fool-ing* contrast, how robust is this contrast phonetically? Recall that, in modular approaches, morphological contrasts can only operate on distinct allophones, corresponding to front and back variants of the GOOSE vowel and/or light and dark /l/. The phonetic prediction that follows from this is that morphologically-conditioned contrasts should involve robust phonetic differences. From the point of view of non-modular theories, on the other hand, we expect to find all kinds of effects sizes associated with the distinction between *hula* and *fool-ing*. We test these two possibilities by applying detailed phonetic analysis to the corpus also presented in Authors (2015) and Authors (Under review). A description of the relevant parts of the corpus and the method are in Section 2, followed by a description of acoustic and articulatory measures. The results, presented in Section 3, show that the vowel in morphologically-complex words like *fool-ing* is typically more retracted than it is in comparable monomorphemes (like *hula*), but it is not as back as it is in unsuffixed *fool*. We also find parallel differences affecting the /l/ in the two contexts. Regarding the degree of retraction in *fool-ing*, it can fall virtually anywhere in the space between *hula* and *fool*. In Section 4, we argue that these gradient and speaker-specific findings are problematic for an analysis which predicts a small set of possibilities, excluding morphologisation that had not been preceded by an abrupt allophonic split. Section 5 concludes.

2 Method

2.1 Stimuli

Our test items included /u:/ followed by /l/ in three different conditions: 1) morpheme-internal, e.g. *hula* 2) morpheme-final word-internal (henceforth, ‘morpheme-final’), e.g. *fool-ing* 3) word-final pre-consonantal, e.g. *fool#(five)*, where *five* is part of the carrier phrase. We used the word-final pre-consonantal context in order to elicit instances of extreme /l/-darkening, and we were interested in seeing the whole /l/-darkening spectrum. The crucial distinction for our research question, however, is that between morpheme-internal and morpheme-final context. Three different lexical items were used per every combination of vowel and condition. Non-lingual consonants, such as labials or /h/, were preferred preceding the /u:l/ sequence. If, due to lexical restrictions, lingual consonants had to be used, they were counterbalanced across the set. Lexical items with yod-insertion before /u:/, such as *mule*, were not used. A full list of test items is in Table 1. The test items were embedded within a fixed carrier phrase: *Say X five times*. In addition to the test stimuli presented above, we also collected data on corner vowels /i:/ and /a:/, data on word final /u:l/ followed by vowels, on the realisation of /ʊ/ (lexical set of FOOT) in different morphosyntactic conditions, as well data on /u:/ and /ʊ/ followed by coronals.

Table 1: Experimental stimuli

a. Morpheme-internal	<i>hula</i> <i>Pooley</i> <i>goulash</i>
b. Morpheme-final	<i>fool-ing</i> <i>pool-ing</i> <i>school-ing</i>
c. Word-final (pre-consonantal)	<i>fool#(five)</i> <i>pool#(five)</i> <i>school#(five)</i>

Given the requirements for the morphosyntactic composition of the stimuli, coupled

with constraints on the initial consonant, we were unable to control for some other factors in our design. Firstly, there is some variation in the vowel following /l/; it was /i/, /ə/, or /æ/ in the morpheme-internal condition, whereas in the morpheme-final condition, /l/ was always followed by /ɪ/. Secondly, we were unable to control for lexical frequency, a factor which has previously been shown to affect /u:/- and /ɪ/-related changes (Sóskuthy et al. 2015 on /GOOSE/-fronting in Derby English, Lin et al. 2014 on /ɪ/-vocalisation in American English). We acknowledge this issue here, and discuss how it might have affected our results in Section 4.

2.2 Speakers

The corpus contains data from 20 speakers: 10 older (3 males 49-66, mean=56, 7 females 45-62, mean=55), and 10 younger (3 males 21-28, mean=25, 7 females 20-25, mean=22). They had all been born and had grown up in the South of England or the English Midlands, and self-identified as speakers of Southern British English. They were not aware of the purpose of the experiment. They were compensated for their participation.

2.3 Procedure

Time-synchronised articulatory and audio data were collected in the experiment. Tongue movement data were captured using a high-speed Sonix RP ultrasound system (Frame Rate = 121.5 fps, Scanlines = 63, Pixels per Scanline = 412, Field of Vision = 134.9°, Pixel offset = 51, Depth = 80 mm). Details concerning this ultrasound system are described in Wrench & Scobbie (2016). The ultrasonic probe was positioned under the participant’s chin and stabilised using a headset (Articulate Instruments Ltd, 2008). The audio data were captured using a lavalier Audio-Technica AT803 condenser microphone. The audio data were sampled at 22 kHz. The stimuli were presented to the participants on a computer screen, one at a time by Articulate Assistant Advanced™ (AAA) version 2.16, which also controlled the synchronisation and data capture. The

order of the stimuli was randomised for each participant. Each participant read four repetitions of the experimental material (four blocks in the same order, with a pause following the second block). In addition, each participant was recorded swallowing water, in order to image the hard palate, and biting on a piece of plastic (a bite plate) while pushing the tongue up to make contact, in order to image the occlusal plane (see Epstein & Stone 2005, Scobbie et al. 2011, and Miller 2016 for description of methods for imaging the hard palate and the occlusal plane with ultrasound). The occlusal plane images were central to the articulatory analysis, as described in Section 2.5.4. The palate traces, on the other hand, were only used in visual exploration of the data (see Figure 5 for an example).

2.4 Segmentation

The acoustic data were exported as wav files from AAA, and automatically segmented using the University of Pennsylvania Forced Aligner (FAVE, Rosenfelder et al. 2011). The automatic segmentation was then hand-corrected by the first author in Praat version 5.3 (Boersma & Weenink, 2009). The annotations were used for duration measurements and formant tracking in Praat, and the boundaries were imported back into AAA for the articulatory analysis. For the purpose of our analysis, we were mainly interested in extracting the initial and the final boundary of the /u:l/ sequence. As these sequences were preceded by obstruents, the automatic segmentation was generally robust in finding the beginning of /u:l/. The boundary between /l/ and the following vowel, for instance in words like *pooling* was typically characterised by relatively abrupt intensity transitions (see Figure 3) that were consistently identified by the software. The boundary between the vowel and the following /l/, on the other hand, was difficult to determine reliably, especially in the word-final contexts, like *pool* (see Figure 4). Since no reliable segmentation strategy could be established to separate the vowel from the /l/, we did not attempt such segmentation. Instead, we proceed in our analysis to consider the /u:l/ sequences as a whole. This is done by a combination of dynamic formant analysis and static acoustic and articulatory analysis which relies on selected

landmarks that were consistently identifiable to select individual points for comparison across tokens.

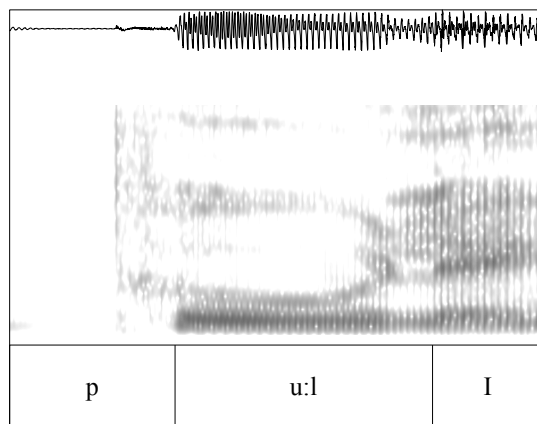


Figure 3: Example segmentation of pre-vocalic /l/

2.5 Analysis

2.5.1 Dynamic formant analysis

For each test item, we extracted duration measurements for the /u:l/ sequence, based on the segmentation procedure described above in Section 2.4. Although we do not expect duration to interact directly with either GOOSE-fronting or /l/-darkening, we included the duration measurements, since Sugahara & Turk (2009) find that duration of stem-final rhymes in morphologically-complex words in English is increased compared to the same segmental sequences in monomorphemic words (see Section 1 above). Therefore, acoustic duration is a phonetic dimension that could be potentially affected by morphological structure. We also measured the first three formants, in Hz, throughout the

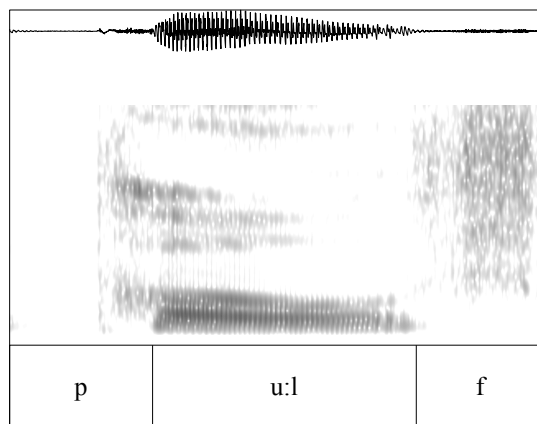


Figure 4: Example segmentation of word-final pre-consonantal /l/

/u:l/ sequence at 10% intervals, using a modified version of the Praat script by Remijsen (2004). The script extracts the first three formants based on the Burg algorithm in Praat, and displays the following for each individual token: the spectrogram, formant tracks, the rounded F1 and F2 values, the Long-term average spectrum (Ltas) and the LPC spectrum. We used the display of spectrogram and overlaid formant tracks to optimise formant settings for each individual speaker. We further eyeballed the displays produced by the script for each token throughout the formant extraction procedure to ensure that tracking accuracy was sustained. We analysed the formant contours in normalised time depending on the morpho-syntactic condition, using Smoothing Spline Analysis of Variance (SS-ANOVA, Gu 2013; Davidson 2006; Docherty et al. 2015), implemented using the `tongue_ssanova.r` functions written by Mielke, based on the `gss` package (Mielke, 2013; Gu, 2014). This method returns mean smoothed formant contours for each condition, along with 95% Bayesian confidence intervals. Non-overlapping confidence intervals are taken as indicating significant differences between conditions.

We compared the formants across the three morpho-syntactic condition, pooling the test items within each condition.

2.5.2 Static formant analysis

While the SS-ANOVAs allow us to capture information concerning the dynamic nature of contrast throughout the /u:l/ sequence, the disadvantage of this analysis is that effect sizes are not easily quantifiable. Therefore, we followed up the dynamic analysis with static measurements of formants at representative time points. From the continuous measurements of the formant tracks, we selected two (for each token): 1) the acoustic onset of the vowel, and 2) the trough of the F2 track. We chose to analyse formants at the acoustic onset of the vowel, in order to investigate whether morpho-syntactic contrasts are observable already from the start of the vowel. The F2 dip was chosen as representative of the /l/ (in its dorsal part). We typically find that the F2 would drop over the initial part of /u:l/, and then rise again (see Figure 3 for an example). We identified such a dip, whenever present, using an automated function for finding peaks (or valleys) in the *pracma* package in R (Borchers, 2015). Where no clear dip could be found, we measured the minimum F2.

We analysed the F2 static measurements using mixed-effects linear regression implemented using the *lme4* package (Bates & Maechler, 2009) in R version 3.02 (R Development Core Team, 2005). The modelling procedure and the final model structure are presented in Section 3.2 along with the results.

2.5.3 Acoustic duration

Similarly to the formant measurements, we analysed acoustic duration of /u:l/ using mixed-effects linear regression. Duration was measured in ms, and we modelled the square root of the measurements to correct for the right-skewness of the distribution.

2.5.4 Articulatory measures

For the articulatory part of our analyses, we traced the mid-sagittal tongue contour in the ultrasound image at the following two time points: at the acoustic onset of the vowel, and at the point of maximal retraction of the tongue dorsum in the occlusal plane during the acoustic /u:l/ sequence. Example ultrasound frames illustrating the position of the tongue at the two points for a single token of *fooling* pronounced by speaker OF1 are in Figure 5. Our choice of representative time points for articulatory analysis is analogous to the static analysis of F2 presented in Section 3.2 above. The point of maximum dorsal retraction was selected to represent the dorsal constriction in the production of /l/, as observed in previous articulatory research on English /l/ (Scobbie & Pouplier, 2010; Sproat & Fujimura, 1993). The magnitude of the dorsal gesture has previously been analysed as a correlate of /l/-darkening by Lee-Kim et al. (2013). Lee-Kim et al. (2013) observe that, for speakers of American English, tongue position at the dorsal maximum varies as a function of morphological structure. They find increased dorsal retraction before a morphological boundary (e.g. *cool-est*), compared to /l/ at the beginning of a morpheme (e.g. *coup-less*).

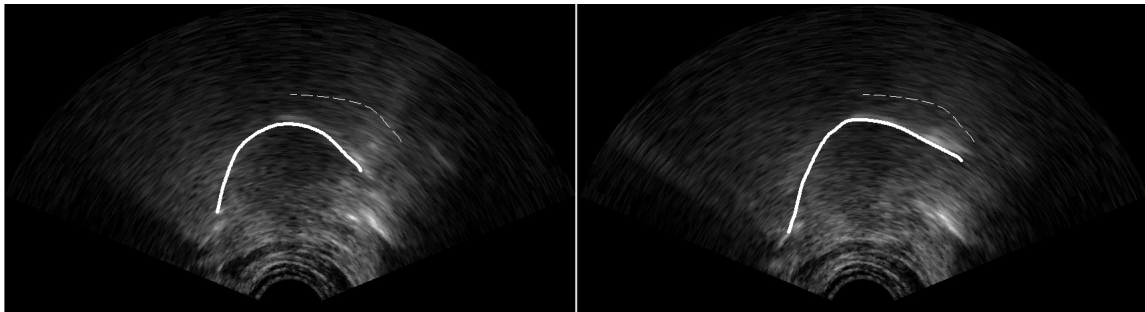


Figure 5: Ultrasound frames showing tongue position tracings for an example token (*fooling*) pronounced by Speaker OF1 at the acoustic onset of the vowel (left), and at the point of maximal dorsal retraction (right). The dashed line shows the palate outline. Note that these images are not rotated in the occlusal plane.

For the two selected time points, we used AAA to semi-automatically trace a smoothed contour from the underside of the bright white areas of the image corresponding to the surface of the tongue, and we subsequently extracted the coordinates

of the tracings. We rotated the extracted coordinates on the occlusal plane for each speaker, to normalise the horizontal and vertical orientation of the tongue in the mid-sagittal plane in a consistent cross-speaker way based on this articulatory landmark. We then compared the mean coordinate values for each morphosyntactic condition, using SS-ANOVA (see Section 3.1). Just like in the previous analyses, we pooled different lexical items in each condition for the analysis. We followed Mielke (2015) in calculating the tongue SS-ANOVA using polar coordinates, which provides more accurate calculation of confidence intervals, especially when the comparisons concern the tongue root. The data from three speakers (OM2, OF5, and YM1) did not allow us to trace any tongue contours, due to poor quality of the ultrasound image. For these speakers, we could see the tongue contour for parts of the recording, but the image was less clear for high vowels, due to increased distance from the probe. We could have achieved clearer image with a low-frequency probe, but such a probe is not as yet available for the Ultrasonix system we use.³ There were further five speakers (OF7, OM1, YF4, YF9, and OM2), for whom we could not confidently find the time point of maximal dorsal retraction. While the quality of tongue image for those speakers was, on average, acceptable, the tongue root was not always clearly visible where the orientation of the tongue surface was nearly parallel to the beams coming from the probe, making it difficult to determine the point of maximal dorsal retraction. In Section 3.4, we only consider articulatory data from the subset of speakers for whom we were able to confidently trace the tongue contour accurately at the relevant time points (17 speakers for the acoustic onset of the vowel, 12 speakers for the /l/).

In order to quantify the relative articulatory differences between *hula* and *fool-ing*, we calculated the area between the curves representing average tongue shapes for *hula* and *fool-ing* (for each speaker). The calculations were done using the R package *flux* (Jurasinski et al., 2014). Because the area measurements are not easy to interpret

³A reviewer asks about the benefits of using a high-speed ultrasound system for a static articulatory analysis. Increased temporal resolution allows us to select the relevant time point for analysis more consistently, reducing the variance in the data, and thus allowing for more precise spatial comparisons. The benefits are similar to measuring acoustic formants at a single time point, which is based on continuous formant values extracted at each cycle.

on their own, we report them along with the area between the curves representing *hula* and *fool* (Figure 6). In calculating the distance, in order to provide a normalised, articulation-based measure, we only considered the part of the curves above the occlusal plane, which corresponds to 0 on the y-axis in our plots. We needed a standardised cut-off point, because whole tongue contours are usually not comparable. This is partly because individuals vary in the size of the tongue, and mainly because ultrasound tongue surface contours are images of an unknown proportion of the whole tongue. For many speakers, hyoid and jaw shadows obscure part of the tongue image, and as a result, individual extracted tongue contours differ in length. Using the occlusal plane as the cut-off point provides a way of dealing with this issue, although some information about the tongue root is lost in this way.

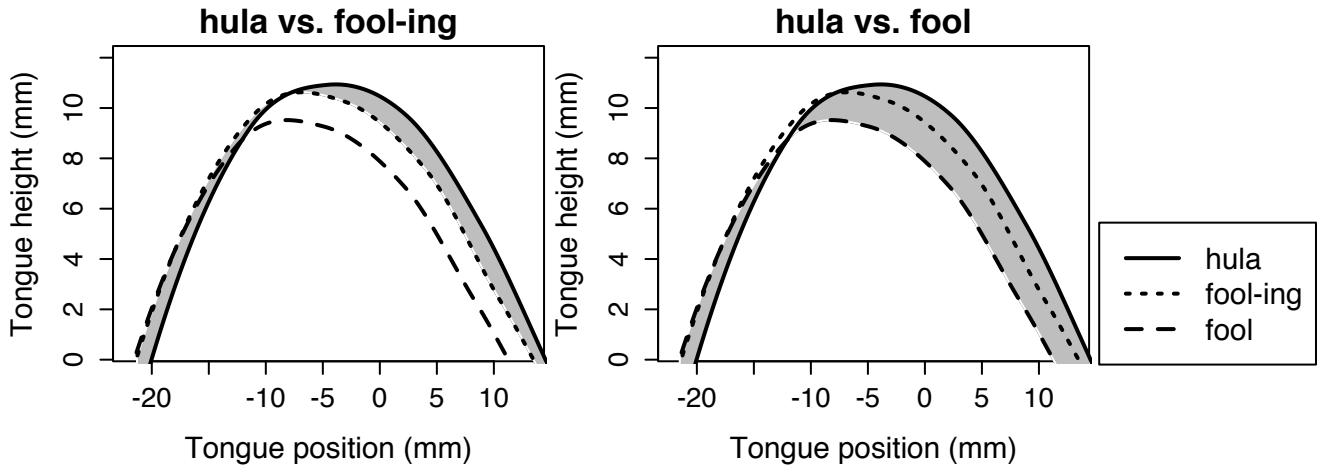


Figure 6: Area between curves representing averaged tongue contours for *hula* and *fool-ing* (left), and *hula* and *fool* (right).

3 Results

3.1 Dynamic formant analysis results

The dynamic formant analysis shows that the main acoustic parameter systematically affected by morphosyntactic condition is F2. The observed relevance of F2 is consis-

tent with the impressionistic descriptions that the *hula*~*fool-ing* distinction involves (non) fronting of the vowel and darkening of the /l/. A relatively more back variant of the GOOSE vowel is expected in *fool-ing*, accompanied by a degree of /l/-darkening. Both of these are expected to involve F2 lowering, based on previous acoustic studies on GOOSE-fronting (Bauer, 1985; Hawkins & Midgley, 2005; McDougall & Nolan, 2007; Harrington, 2007; Harrington et al., 2008; Chládková & Hamann, 2011) and /l/-darkening (Ladefoged & Maddieson, 1996; Carter, 2002, 2003; Carter & Local, 2007; Hawkins & Nguyen, 2004).

Figure 7 shows the results of SS-ANOVA analysis of normalised F2 contours within each individual speaker, as a function of morpho-syntactic condition. Note that, in this plot, and in subsequent ones, we use *hula*, *fool-ing* and *fool* as category labels denoting the three experimental conditions: morpheme-internal, morpheme-final, and word-final pre-consonantal. The means include data pooled across different lexical items in each morphosyntactic condition.

Most speakers in our pool, excepting OF4 and OM4, produced an F2 distinction between *hula* and *fool-ing*. Whenever the distinction appeared, it was realised in the expected direction, i.e. *hula* showed a relatively higher F2, indicating more front vowel pronunciation and/or relatively lighter /l/. Furthermore, all speakers produced a contrast between *hula* and *fool*, once again showing F2 lowering in *fool*. For most speakers, *fool-ing* did not clearly pattern with *fool*, but it formed an intermediate category between *fool* and *hula*. Speaker YM1 was an exception in this respect, as he had two clear acoustic categories: one for *fool* and *fool-ing*, and a separate one for *hula*.

Impressionistically, in contrast to the prediction made by the modular approach, the *hula*~*fool-ing* contrast is not necessarily associated with big effect sizes. If the emergence of the contrast between those two categories involves an allophonic shift in *fool-ing*, we would expect a robust difference between /u:/ and/or /l/ in those two cases. In contrast, our data indicate that the size of this difference is extremely variable, ranging from barely significant differences in the initial (vocalic) portion (OF3, OF5), through a range of intermediate realisations (e.g. OF6, YF7) to large differences (YF6,

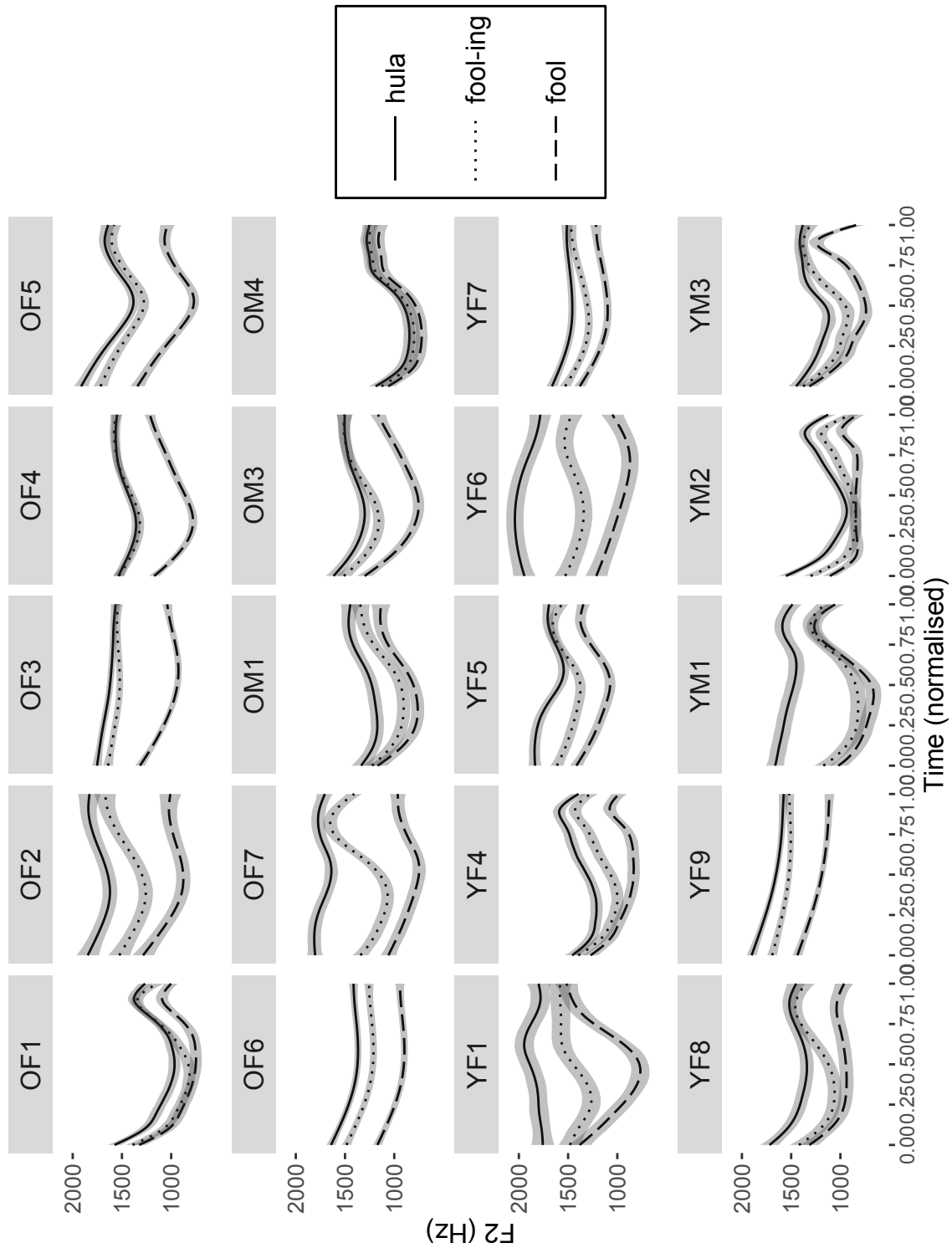


Figure 7: Results of individual speaker SS ANOVA analyses of time-normalised F2, depending on the morpho-syntactic context. Individual speaker codes include information about age group (O=older, Y=younger) and sex (F/M).

YM1).

We did not find evidence in our data of the *hula*~*fool-ing* contrast systematically affecting F1 or F3 (although there were sometimes clear F1 and/or F3 differences between word-final *fool* and word-medial *hula/fool-ing*). Some speakers produced a small difference between *hula* and *fool-ing* along those dimensions, but for most speakers, there was no difference. The results on the SS-ANOVA analysis for F1 and F3 are in the Appendix.

3.2 Static formant analysis results

The SS-ANOVA comparisons of F2 tracks presented in Section 3.1 above suggests that the distance between *hula* and *fool-ing* along the F2 dimension is highly variable. In order to quantify this observation, we fitted linear mixed-effects regression models to F2 measurements at two time points. This analysis allows us to investigate individual variation with respect to effects sizes in a more formal way, and it also allows us to compare the two age groups we are investigating.

We present two models, one for F2 at the acoustic onset, and one for F2 at the F2 track dip. The F2 (measured in Hz) was the dependent variable. The modelling procedure was as follows. In the initial model, we included the following fixed predictors: age (older vs. younger), sex (male vs. female) and morpho-syntactic context (labelled *hula*, *fool-ing*, *fool*). We also included F1 as a co-variate. The random structure included random intercepts for speaker and item, as well as random slopes for morpho-syntactic context within speaker. As the next step, we used the log-likelihood comparison (Baayen, 2008) to test for interactions between age and morpho-syntactic context, sex and morpho-syntactic context, and age, sex and morpho-syntactic context. Non-significant interactions were not retained. For the final model, we verified whether it provided a significantly better fit compared to a model with the same fixed-effects structure, and random intercepts only, and compared to a model with no random intercept for item.

The final model we selected for the onset F2, had main effects of age, sex, F1 and

morpho-syntactic context, as well as an interaction between age and morpho-syntactic context (see Table 2). F1 and F2 were positively correlated, and male speakers had, on average, lower F2 than female speakers. The F2 was, on average, highest in *hula* followed by *fool-ing* and *fool*. In all three conditions, younger speakers had higher F2 than older speakers, but the apparent time difference was largest for *hula*. The interaction is plotted in Figure 8. It suggests an effect of /u:/-fronting in apparent time that is relatively largest in the transparent monomorpheme condition. Note, however, that there is a considerable degree of individual variation in this respect, as indicated by relatively high Standard Error values (Table 2). This is further corroborated by the significant effect of context within speaker, as indicated by a significant improvement in fit for a model with a random slope, compared to a model with random intercepts only ($\chi^2 = 66.69$, $df = 5$, $p < .001$). Thus, we find individual variation that is not explained by the age differences. The random item effects were also significant ($\chi^2 = 153.23$, $df = 6$, $p < .001$).

Table 2: Summary of fixed effects of a linear mixed-effects regression model predicting the F2 (in Hz) measured at the acoustic onset of the vowel. The intercept corresponds to *hula* pronounced by an older female speaker.

Term	Level	Estimate	Std. Error	<i>t</i> -value
(Intercept)		1537.08	91.53	16.79
F1		0.23	0.10	2.31
Age	younger	104.28	82.32	1.27
Context	<i>fool-ing</i>	-184.60	92.53	-2.00
Context	<i>fool</i>	-372.25	100.22	-3.71
Sex	M	-187.39	47.07	-3.98
Age:Context	younger <i>fool-ing</i>	-96.02	62.69	-1.53
Age:Context	younger <i>fool</i>	-53.82	82.76	-0.65

We further explored the inter-speaker variation with respect to difference between *hula* and *fool-ing*. We extracted the mean distance between *hula* and *fool-ing* for each speaker, based on the random slopes in the model. We then examined the distribution of the differences, shown in Figure 9. The bin-width was 10 Hz. Each 10Hz panel corresponds to a single speaker. Increased bin width or bin height indicates a concentration

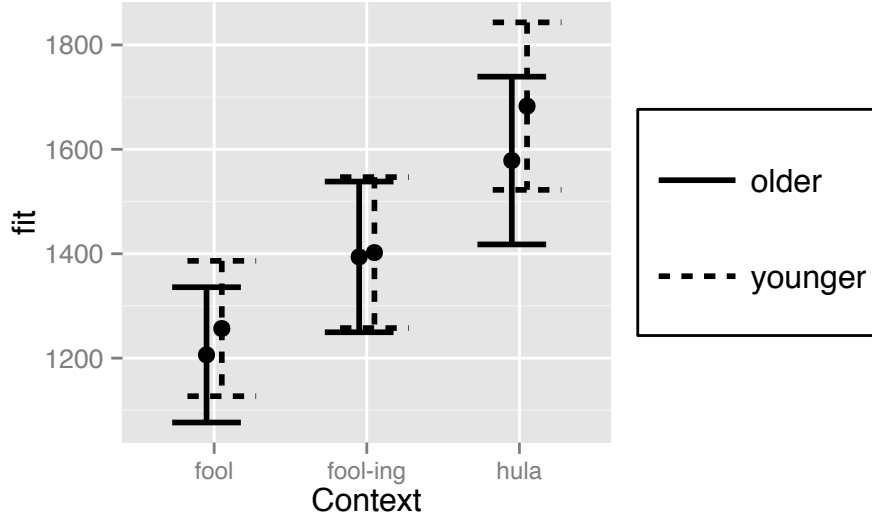


Figure 8: Interaction between morpho-syntactic condition and age in a linear mixed-effects regression model predicting the F2 (in Hz) measured at the acoustic onset of the vowel.

of speakers who show a difference within a specific range.

As we can see in Figure 9, the distribution of the differences is fairly continuous. Three speakers (these were OF7, YF6 and YM1) were quite distinct in having a relatively large difference between *hula* and *fool-ing*. However, we see no similar break at the other end of the continuum which includes speakers for whom the *hula* and *fool-ing* differences was significant and those for whom it was not. A significant difference seems to emerge at some point of a steady trend, and it is not associated with a categorical break.

We repeated the same modelling procedure for the value of the F2 dip (we can interpret this analysis as comparing F2 values at the acoustic /l/-darkening target). The best fitting model had a main effect of F1, sex, age and morpho-syntactic context in its fixed part, but no significant interactions. The summary of the fixed part of the model is in Table 3. F2 was higher in younger speakers compared to older speakers, and higher in females, compared to males. As far as the morpho-syntactic context was concerned, F2 was significantly higher in *hula* than in *fool-ing* or *fool* (the mean difference between *fool-ing* and *fool* was -268.56 Hz, $SE = 67.65$, $t = -3.97$, as established

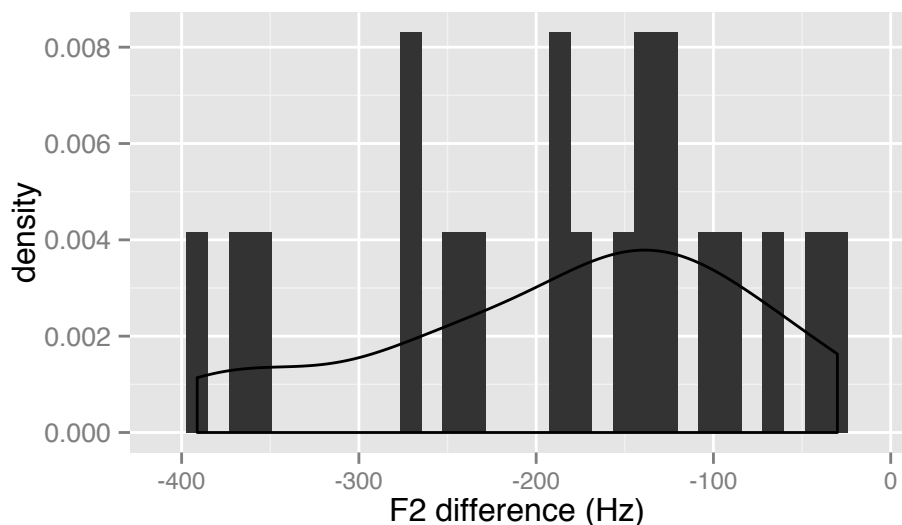


Figure 9: Histogram and density plot representing the distribution of mean F2 differences (measured at the vowel onset) between *hula* and *fool-ing* for each speaker

by re-fitting a model after re-levelling the context factor). However, there was no significant interaction between age and morpho-syntactic context. This is different from the model of onset F2, and it suggests the absence of apparent-time changes in the acoustic /l/-darkening target. While there were no systematic age differences, there was individual variation, confirmed by the significant improvement in fit shown by the comparison between a model with random slopes and the model with random intercepts only ($\chi^2 = 285.5$, $df = 5$, $p. < .001$). The inclusion of a random intercept for item improved the model significantly ($\chi^2 = 321.91$, $df = 6$, $p. < .001$).

Table 3: Summary of fixed effects of a linear mixed-effects regression model predicting the F2 (in Hz) measured at the F2 dip. The intercept corresponds to *hula* pronounced by an older female speaker.

Term	Level	Estimate	Std. Error	<i>t</i> -value
(Intercept)		906.47	98.74	9.18
F1		1.18	0.18	6.46
Age	younger	84.40	40.00	2.11
Context	<i>fool-ing</i>	-249.17	70.88	-3.52
Context	<i>fool</i>	-517.74	81.66	-6.34
Sex	M	-110.67	43.63	-2.54

In order to explore the individual variation, we examined the distribution of the difference between *hula* and *fool-ing* for each speaker (based on the model output). The distribution is plotted in Figure 10. The bin width was 20 Hz. 14 out of the 20 speakers cluster within the range of 50 to 250 Hz of difference. This includes speakers like (OF4 and OF5) who make no significant difference between *hula* and *fool-ing* at the F2 dip (compare Figure 7), and speakers like YF7 or OF5, who do. At the other end of the distribution, we see speakers with more categorical effects. The most extreme differences were produced by speakers YF6, YM1 and OF7.

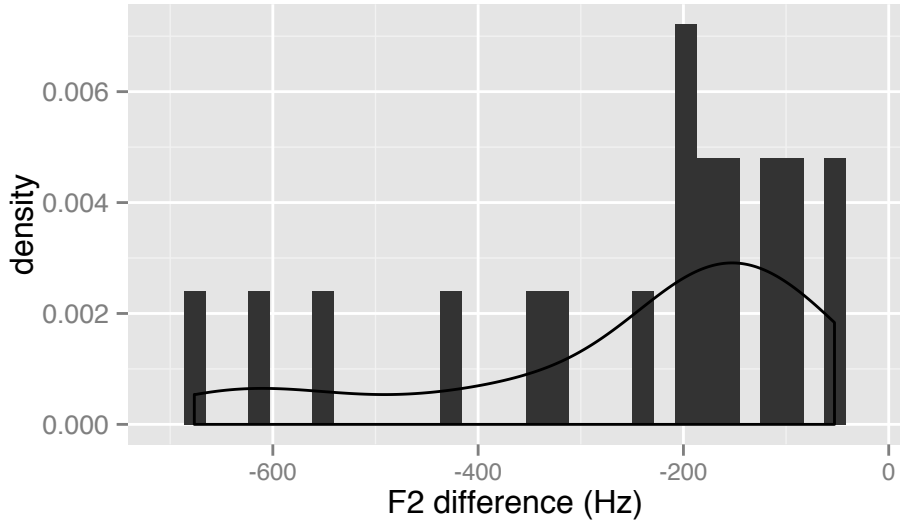


Figure 10: Histogram and density plot representing the distribution of mean F2 differences (measured at the F2 dip) between *hula* and *fool-ing* for each speaker

3.3 Acoustic duration results

In modelling acoustic duration, we used the following model selection procedure. We initially fitted a model with three main predictors: age (older vs. younger), morpho-syntactic context (*hula*, *fool-ing* or *fool*) and sex (male vs. female). The random predictors included speaker, item and morpho-syntactic context within speaker. We then tested for an interaction between age and context, to verify whether some contexts vary more than others in apparent time. This interaction was not significant ($\chi^2 = 2.25$,

$df = 2, p = 0.32$), and it was not retained. We then considered whether the inclusion of a random slope for condition within speaker improved the model fit. This was confirmed ($\chi^2 = 187.86, df = 5, p < .001$). Based on this procedure, we selected the initial model as the best-fitting one. The random intercept for item was significant ($\chi^2 = 8907, df = 1, p < 0.01$). The model summary is in Table 4.

Table 4: Summary of fixed effects of a linear mixed-effects regression model predicting the square root of /u:l/ duration. The intercept corresponds to *hula* pronounced by an older female speaker.

Term	Level	Estimate	Std. Error	<i>t</i> -value
(Intercept)		15.28	0.44	35.07
Age	younger	-0.92	0.44	-2.10
Context	<i>fool-ing</i>	0.10	0.28	0.36
Context	<i>fool</i>	3.90	0.48	8.06
Sex	M	-0.75	0.48	-1.55

Older speakers had an overall increased duration compared to younger speakers, but this did not vary across morpho-syntactic contexts. There was a main effect of morpho-syntactic context, with lengthening in *fool*, compared to *hula* or *fool-ing*. There was, however, no significant difference in duration between morpheme-internal and morpheme-final (*hula~fool-ing*) context. In contrast, we find increased duration in the word-final *fool#* compared to the two word-medial environments. This lengthening effect is expected, considering the shape of our carrier phrase (*‘Say fool five times’*). A relatively strong prosodic boundary may occur following *fool*, which is likely to involve segmental lengthening (Byrd, 2000; Byrd & Saltzman, 2003; Fougeron & Keating, 1997; Turk & Shattuck-Hufnagel, 2000). We also find individual variation with respect to how morpho-syntactic context affected duration. Further analysis of random slopes showed that this variation mainly involved the difference in duration between *fool* vs. *hula* or *fool-ing*. We interpret this result as conditioned by individual difference with respect to how much prosodic boundary marking was present in *fool*.

3.4 Articulatory results

This section presents the results of SS-ANOVA comparisons of tongue contour tracings at two selected time points for individual speakers. According to this comparison, all speakers except two (YF9 and OF3) had a significant contrast between *hula* and *fool-ing* at the acoustic onset of the vowel. Figure 11 shows the results of SS-ANOVA for four example speakers at the vowel onset. For all of them the tongue position is significantly fronted in *hula* compared to *fool-ing*. For speakers OF4, OM4 and YF6, this difference in position mainly affects tongue root and dorsum, while speaker YF8 shows a significant difference in the pre-dorsum. Notice that speakers OF4 and OM4 showed no acoustic difference between *hula* and *fool-ing* at the vowel onset or indeed at any other time point (Figure 7). Nevertheless, they make an articulatory difference between these two conditions. While the size of the difference is extremely small, we can see fronting differences in the tongue root in the predicted direction. The effect sizes for other speakers vary considerably. Speaker YF8 represents the most common pattern, which involves a fairly small, but significant, three-way distinction, with increasing fronting in the direction $fool < fool-ing < hula$. The data from speaker YF6 illustrate that the difference in tongue shape between *hula* and *fool-ing* can also be quite large.

In the interest of space, we do not provide individual SS-ANOVA result plots for all individual speakers. Instead, we summarise the distances between *hula* and *fool-ing* based on area between the two curves, as described in Section 2.5.4. We also report the distances between *hula* and *fool*, to make the area measurements more interpretable, although the reader should bear in mind that the *hula*~*fool* difference could potentially vary between speakers, depending on individual differences in the prosodic realisation of *fool*. The measurements are in Table 5. We find a big spread of values for the *hula*~*fool-ing* distance with a concentration in the centre of the scale. Typically, the tongue shapes for *fool-ing* were roughly equidistant from both *hula* and *fool*. However, we also find some very small values which nevertheless correspond to significant *hula*~*fool-ing* differences.

Figure 12 show the results of SS-ANOVA for the same four example speakers whose

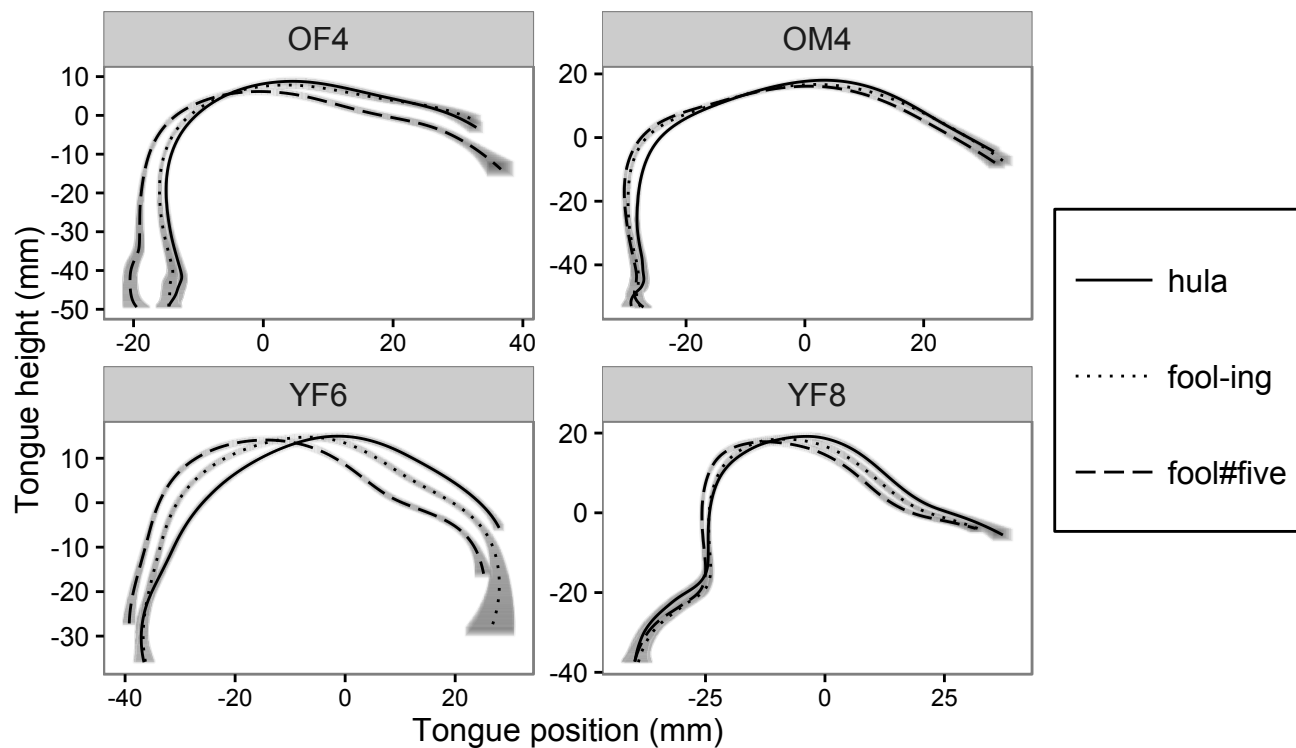


Figure 11: Results of SS-ANOVA comparison of tongue contour as a function of morphosyntactic condition for speakers OF4, OM4, YF6 and YF8, measured at the acoustic onset of the vowel. The tongue tip is on the right. The images have been rotated, so that the x-axis is in the occlusal plane, but they are not anatomically accurate, due to coordinate transformation and scaling.

Table 5: Summary of areas (in mm²) between the curves representing average tongue contour tracings for /u:/ extracted at the acoustic onset of the vowel and fitted by SS-ANOVA for individual speakers. Non-significant differences between curves are indicated by shading.

Speaker	<i>hula</i> ~ <i>fool-ing</i> distance	<i>hula</i> ~ <i>fool</i> distance
YF9	19.68	146.49
OF4	25.30	85.69
OF1	32.86	61.19
OF6	34.56	91.52
OM4	43.98	92.89
YM3	46.35	101.61
YF5	54.72	132.58
YF7	70.83	94.13
OF3	73.88	182.26
YF1	74.44	147.55
YF8	82.92	163.79
YF4	89.95	233.33
OM1	93.78	98.43
OF7	126.18	177.38
YF6	149.51	215.46
OF2	149.78	226.02
YM2	152.49	169.26

vowel onset data we saw in Figure 11. At the /l/ dorsal target, Speakers OF4 and OM4 showed a marginally significant difference in the tongue root between *hula* and *fool-ing*. In comparison, they have clear dorsal retraction in *fool*. For speaker YF6, the tongue position in *fool-ing* is approximately half-way between *hula* and *fool*, with a clearly significant three-way distinction. For speaker YF8, *fool-ing* patterns with *hula* as far as tongue root is concerned, but, at the same time, *fool-ing* shows tongue dorsum raising and retraction that are similar to *fool*. A summary of the values of areas between curves representing *hula* and *fool-ing* and *hula* and *fool* is in Table 6. All speakers had a significant contrast in articulation between *hula* and *fool-ing* at the dorsal maximum. The relative distances between *hula*~*fool-ing* and *hula*~*fool* varied between speakers. Depending on the speaker, *fool-ing* might be closer to *hula* or to *fool*.

Since our analysis of /l/ focuses on the point of maximum dorsal retraction, it does not allow us to assess systematically what happens with the tongue tip. Whereas

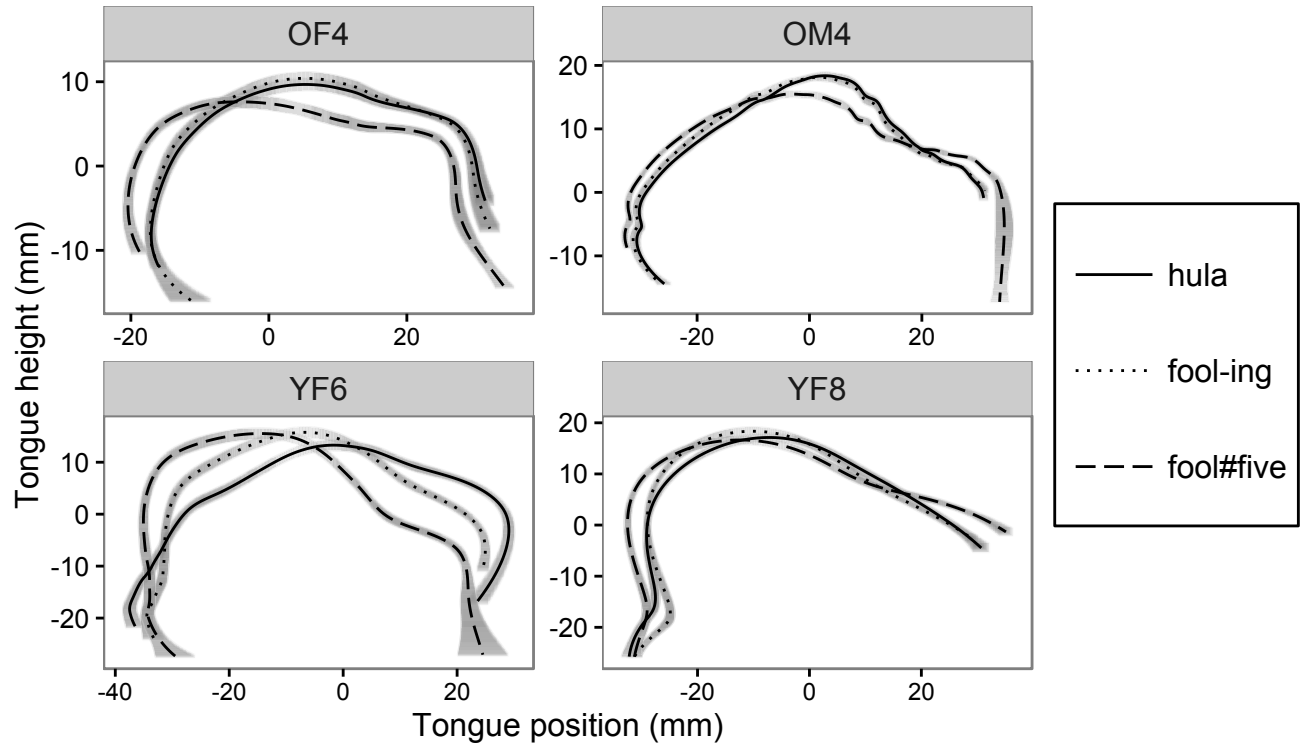


Figure 12: Results of SS-ANOVA comparison of tongue contour as a function of morphosyntactic condition for speakers OF4, OM4, YF6 and YF8, measured at the point of maximal dorsal retraction. The tongue tip is on the right.

Table 6: Summary of areas (in mm²) between the curves representing average tongue contour tracings for /l/ extracted at the dorsal maximum and fitted by SS-ANOVA for individual speakers. All differences were significant.

Speaker	<i>hula~fool-ing</i> distance	<i>hula~fool#five</i> distance
OF4	28.53	108.27
OF3	34.78	138.31
OF1	37.65	75.27
OM4	44.36	142.25
OF6	46.89	157.50
YF5	55.40	89.63
YF8	76.40	120.87
YF7	87.05	145.28
YM2	87.98	254.17
YF1	89.47	145.35
OF2	93.46	131.68
YF6	197.38	240.80

/l/-darkening can be measured independently of the tongue tip gesture, we may ask whether the tongue tip gesture is reduced, showing effects of /l/-vocalisation, and /l/-vocalisation could be seen as another distinct stage of /l/-darkening. We do not intend to present a full-fledged analysis, but briefly, an indication of the tongue tip gesture was always clearly visible for the word-medial pre-vocalic /l/, both in morpheme-medial and morpheme-final contexts. Word-finally (in *fool#(five)*), on the other hand, there was variation, and we sometimes see the tongue tip gesture being reduced and/or delayed, especially in younger speakers.

4 Discussion

The main question we are trying to tackle in this work is whether morphologically-conditioned allophones arise in a phonetically gradual or abrupt manner. In the former case, we predict that morphological contrasts initially involve small and not necessarily segmental phonetic differences, and that those phonetic differences increase gradually over time and become associated with categorical segmental allophones. The alternative hypothesis is that morphological contrasts are initially absent, but when they do appear, they involve a clear opposition between two phonetically distinct allophones.

We do not have longitudinal data to test those two diachronic scenarios explicitly. However, the nature of phonetic variation, as conditioned by morpho-syntactic categories in our data, is certainly more consistent with a gradual interpretation. We considered articulatory and acoustic realisation of /l/-darkening and contextual GOOSE-fronting in Southern British English. Those two processes interact with each other and with the morphological boundary following /l/. When intervocalic /l/ is morpheme-final (*fool-ing*), the most retracted frame of /l/ is darker compared to morpheme-internal /l/ (*hula*), and the preceding /u:/ vowel has more retracted tongue position at its onset. Importantly, we have examined the patterns of variation between individuals for both /u:/ and /l/, and we find that individuals vary a lot as far as effect sizes are concerned. For some speakers, the /u:l/ sequences in *hula* and *fool-ing* are very similar indeed, but the small differences we see between them are significant. For some other speakers, the /l/-darkening and vowel retraction in *fool-ing* can be quite advanced, approaching what we find in *fool*. Mostly, we find a whole range of intermediate realisations: the phonetic distance between *hula* and *fool* can vary considerably, and *fool-ing* tends to show an intermediate realisation between those two phonetic categories.

It is extremely challenging to classify all of our participants as having a ‘big’ or ‘small’ phonetic distance between *hula* and *fool-ing* if we want to use this criterion to determine whether the contrast is mediated by an intermediate allophonic representation. What degree of phonetic difference, exactly, is ‘big enough’ to be allophonic? The ‘big’ vs. ‘small’ opposition is well-defined in terms of prototypes, but the boundaries are blurry. However, we believe that at least some speakers in our study show prototypically small contrasts. This includes speakers, like OM4 and OF4, who make minute articulatory distinctions between the two morphological categories, such that there is no clear acoustic difference. The observation that morphological distinctions can be realised in such subtle ways lends support to the idea that morphological effects originate as epiphenomenal, marginal and elusive. In contrast, it is difficult to make the argument that, for speakers such as OM4 and OF4, there are distinct allophones of /u:/ or /l/ (or both) in *hula* and *fool-ing*.

Our argument concerning the more-or-less phonological status of phonetic differences involves the assumption that categorically distinct allophones can indeed be diagnosed phonetically. Allophones are phonologically not contrastive, and they can be defined as contextual variants that are also phonetically distinct. It is most challenging to operationalise such distinctness. We have already mentioned the difficulty of deciding what counts as phonetically distinct or phonetically robust, although such difficulty is not specific to our case, rather it is inherent to the concept of allophony. Some proposals have been made that categorical phonetic behaviour may serve to define what counts as an allophone (Cohn, 2006, 2007; Ladd, 2006; Scobbie, 2007). Turton (2014) makes a case for categoricity in /l/-darkening, arguing that different realisations of /l/ may sometimes cluster around specific category peaks, and that clear discontinuities in the phonetic realisation of /l/ can be taken as category boundaries. Turton analyses the articulatory characteristics of /l/ at its acoustic midpoint, in a variety of phonological contexts, observing that while there may be many /l/-variants within an individual speaker, some of them cluster together, forming families of closely related tongue shapes. However, if we extend this diagnostic of allophony to our data, we do not always find a clear category boundary between monomorphemic *hula* and morphologically complex *fool-ing*. For some of our speakers, there is no evidence of category boundaries, yet there is morphological contrast. Cases like this challenge the idea that sound change follows a clear and empirically tractable progression through distinct levels of grammar.

If we assume that distinct allophones are defined by their phonetic characteristics, we find instances in our data, where morphological distinctions affect gradient phonetic differences, rather than categorical allophones, and therefore the sound change we are considering cannot be modelled in a strictly modular approach. Alternatively, we can reject the idea that phonological categories can indeed be identified based on phonetic evidence. A solution like this is proposed by Fruehwald (2013), based on evidence from /ai/-raising in Philadelphia English. The /ai/-vowel raises to [ɹi] before voiceless obstruents, e.g. in *write*, but raising is blocked before voiced obstruents, e.g. in *ride*.

However, a parallel difference in /ai/-raising applies in *writer* and *rider*: /ai/-undergoes raising in the former case, but not the latter, even though a /t/-flapping rule obliterates the voicing distinction between underlying /t/ and /d/ in the intervocalic position. Fruehwald (2013) shows that /ai/-raising was sensitive to the *writer*~*rider* contrast from the very onset of the change, before two distinct vowel categories emerged, even though the flapping rule was already in place. Based on this evidence, Fruehwald argues that such behaviour constitutes evidence for the phonological nature of sound change, regardless of whether distinct phonetic categories are involved.

Our data are similar to the Philadelphia /ai/ case in the sense that sound change behaves in a phonologically opaque way from the start, and not only once it gets past a specific phonetic landmark. Fruehwald takes this as an argument that phonologisation of sound change does not have clear phonetic correlates, and he critiques the enterprise of defining allophones based on phonetic criteria, while still embracing a modular distinction between phonology and phonetics. This is a possible interpretation, but it takes away from modularity any unique empirical predictions concerning sound change. A modular model may postulate that morphological distinctions only affect allophones, but without an independent definition of what counts as an allophone, its predictive power becomes limited.

Before we conclude, we must address two potential issues with the experimental design, to address the question that the differences reported really are due to morpho-syntactic structure. As mentioned previously in Section 2.1, we were unable to control for some segmental factors, notably the vowel following the /l/. In the morpheme-final context, the following vowel was always /ɪ/, whereas in the morpheme-internal contexts, the vowel could be /i/, /ə/, or /æ/. We have no reason to believe that this particular set of vowels would trigger tongue root fronting at the onset of the vowel in the morpheme-internal contexts, with the possible exception of /i/. More importantly, our main argument is based on the observation that the same morpho-syntactic manipulation may cause extremely small differences in some speakers, whereas for other speakers, the differences are very robust, and we also see a range of intermediate cases. Since

all the speakers read the same experimental materials, the individual differences are arguably better explained by change in progress than by speaker-specific coarticulatory influences.

As far as lexical frequency is concerned, frequency effects are confounded with morpho-syntactic categories in our study: the morpheme-internal category only contains low-frequency words (*hula*, *Pooley*, *goulash*). This category seems to be ahead, as far as /u:/-fronting is concerned. We ascribe this difference to morpho-syntactic structure, but one could make the argument that low-frequency words lead GOOSE-fronting. We are not aware of any previous experimental work on GOOSE-fronting that would specifically corroborate this scenario. On the contrary, in a study of GOOSE-fronting in Derby English, Sós-kuthy et al. (2015) find that high-frequency words lead the change, although this is in the specific context of words which may be subject to variable yod-dropping. The question of direction in which frequency effects are predicted to develop is an interesting one, in the light of recent findings on the role of frequency in sound change. While it has previously been shown for reductive sound changes that high-frequency words lead the change, in a study of a vowel shift in New Zealand English, Hay et al. (2015) find the opposite frequency effect, where low-frequency words are ahead of the high-frequency ones. Importantly, however, the effect sizes reported by Hay et al. are extremely small. The authors make this explicit, suggesting that their findings were only possible due to an extremely large sample size (80646 tokens, 549 speakers with birth-dates spanning 136 years). We find much larger average apparent-time differences between *hula*- and *fool-ing*-type words, comparing 20 speakers separated by one generation. Even if low-frequency is a factor in the rate of GOOSE-fronting, it is unlikely to be the sole factor responsible for our results. For future research on this topic, it would be worthwhile to attempt to isolate the effect of frequency, by studying the effect of morphological structure on the pronunciation of nonce words, for instance through an experimental paradigm like the one in Mousikou et al. (2015). Meanwhile, our findings on real words provide an important baseline for interpreting such research.

Another potential line of future investigation is a study into potential structure

underlying the individual variation we find in our study. Apart from the general age effects (e.g. higher F2 in *hula* at the vowel onset), we also find differences that are not explained by age, as evidenced by the significant random slope effects in F2 analysis (Section 3.2). While a degree of individual variation is expected when there is change-in-progress, we cannot exclude that more systematic variation may emerge in a sociolinguistically-stratified study of the same phenomenon. Recent results by Baranowski (2014) and Turton & Baranowski (2015) show that, in Manchester English (North-West of England), the degree of GOOSE-fronting before /l/ varies as a function of social class: GOOSE-fronting is strongly limited before coda /l/ for a middle-class group of speakers (similar to the SBE pattern reported here), whereas the working-class group produce fronting across the board. The existence of such interactions between segmental and social factors in other English dialects certainly invites further sociolinguistic investigation into potential context-specific social class differences in SBE and elsewhere.

5 Conclusion

In this paper, we considered an ongoing case of morphologisation in sound change to address the question whether sound change can be sensitive to morphological boundaries at any stage of its development, or whether morphological effects only appear in advanced sound changes, once distinct categories have emerged after the the extent of a phonetic difference has crossed some strict but undefined threshold. The former is predicted by non-modular approaches to grammatical architecture, whereas the latter prediction follows from the models of sound change that assume strict modularity of grammar and a limited set of categories that are used to represent allophones.

Our results on /u:/ followed by /l/ in Southern British English show that the interaction of /l/-darkening and contextual /u:/-fronting leads to the emergence of morphological contrast, and so /u:/ and /l/ are pronounced differently in *hula* and *fool-ing*. We observe such differences in 20 speakers participating in our study, including speakers

who produce extremely subtle differences in their pronunciation of /u:l/, depending on morphology. The relevant differences in articulation may be so small that they do not even yield significant effects in the corresponding acoustic parameters. This suggests that morphological boundaries may affect phenomena that are phonetically continuous and gradient, and not only clear cases of allophony. This finding is straightforwardly accommodated by non-modular approaches, but challenging to the modular ones. While our data can, with some revisions, be modelled within a modular grammar, the revised theory no longer makes any unique empirical predictions concerning the development of sound change.

Acknowledgments

We wish to thank the speakers for participating in our study, Steve Cowen for assistance with the recordings, and Alan Wrench for help with the ultrasound system. We also thank the editor, Marija Tabain, and three anonymous reviewers for their comments on the manuscript. The research reported in this paper was supported by a British Academy Postdoctoral Fellowship PDF/pf130029 to the first author.

References

- Aitken, A. J. (1981). The Scottish Vowel-length Rule. Benskin, M. & M. L. Samuels (eds.), *So Many People, Longages, and Tongues, Edinburgh: Middle English Dialect Project*, pp. 131–157.
- Articulate Instruments Ltd (2008). Ultrasound stabilisation headset users manual, revision 1.4.
- Authors (2015). Velocity measures in ultrasound data. Gestural timing of post-vocalic /l/ in English. *Proceedings of the 18th International Congress on Phonetic Sciences*, URL <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/proceedings.html>.

- Authors (Under review). Whence the fuzziness? Morphologisation of interacting sound changes in Southern British English.
- Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge Univ Press.
- Baranowski, M. (2014). The sociolinguistics of back vowel fronting in Manchester English. Paper presented at Methods in Dialectology XV, Groningen.
- Bates, D. & M. Maechler (2009). *lme4: Linear mixed-effects models using Eigen and Eigen++*. URL <http://CRAN.R-project.org/package=lme4>. Accessed 13 August 2010. R package version 0.999375-32.
- Bauer, L. (1985). Tracing phonetic change in the received pronunciation of British English. *Journal of Phonetics* 13, pp. 61–81.
- Beňuš, Š. (2012). Phonetic variation in Slovak yer and non-yer vowels. *Journal of Phonetics* 40, pp. 535–549.
- Bermúdez-Otero, R. & G. Trousdale (2012). Cycles and continua: on unidirectionality and gradualness in language change. Nevalainen, T. & E. C. Traugott (eds.), *Handbook on the history of English: rethinking and extending approaches and methods*, New York: Oxford University Press, pp. 691–720.
- Boersma, P. & B. Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32, pp. 45–86.
- Boersma, P. & D. Weenink (2009). Praat: doing phonetics by computer [Computer programme]. URL <http://www.praat.org/>. Accessed 15 October 2009. Version 5.1.12.
- Borchers, H. W. (2015). *pracma: Practical Numerical Math Functions*. URL <http://CRAN.R-project.org/package=pracma>. R package version 1.8.6.
- Bybee, J. (2001). *Phonology and Language Use*. Cambridge, MA: Cambridge University Press.

- Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica* 57, pp. 3–16.
- Byrd, D. & E. Saltzman (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31, pp. 149–180.
- Carter, P. (2002). *Structured variation in British English liquids*. Ph.D. thesis, University of York.
- Carter, P. (2003). Extrinsic phonetic interpretation: spectral variation in English liquids. Local, J., R. Ogden & R. Temple (eds.), *Phonetic interpretation: Papers in laboratory phonology VI*, Cambridge: Cambridge University Press, pp. 237–252.
- Carter, P. & J. Local (2007). F2 variation in Newcastle and Leeds English liquid systems. *Journal of the International Phonetic Association* 37, pp. 183–199.
- Chládková, K. & S. Hamann (2011). High vowels in Standard British English: /u/-fronting does not result in merger. *Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong*, pp. 476–479.
- Chládková, K., S. Hamann, D. Williams & S. Hellmuth (2016). F2 slope as a perceptual cue for the front-back contrast in Standard Southern British English. Published online in *Language and Speech*.
- Cho, T. & P. Keating (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29, pp. 155–190.
- Cohn, A. C. (2006). Is there gradient phonology? Fanselow, G., C. Féry, M. Schlesewsky & R. Vogel (eds.), *Gradience in Grammar: Generative Perspective*, Oxford: OUP, pp. 25–44.
- Cohn, A. C. (2007). Phonetics in phonology and phonology in phonetics. *Working Papers of the Cornell Phonetics Laboratory* 16, pp. 1–31.

- Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America* 120, pp. 407–415.
- Docherty, G., S. Gonzalez & N. Mitchell (2015). Static vs. dynamic perspectives on the realization of vowel nuclei in West Australian English. *Proceedings of the 18th International Congress on Phonetic Sciences*, URL <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/proceedings.html>.
- Epstein, M. A. & M. Stone (2005). The tongue stops here: Ultrasound imaging of the palate. *The Journal of the Acoustical Society of America* 118, pp. 2128–2131.
- Fougeron, C. & P. Keating (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America* 101, pp. 3728–3740.
- Fruehwald, J. (2013). *The phonological influence on phonetic change*. Ph.D. thesis, University of Pennsylvania.
- Gu, C. (2013). *Smoothing spline ANOVA models*, vol. 297. Springer Science & Business Media.
- Gu, C. (2014). Smoothing Spline ANOVA Models: R package gss. *Journal of Statistical Software* 58:5, pp. 1–25, URL <http://www.jstatsoft.org/v58/i05/>.
- Harrington, J. (2007). Evidence for a relationship between synchronic variability and diachronic change in the Queen’s annual Christmas broadcasts. Cole, J. & J. I. Hualde (eds.), *Laboratory phonology*, Berlin: Mouton de Gruyter, vol. 9, pp. 125–143.
- Harrington, J., F. Kleber & U. Reubold (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *The Journal of the Acoustical Society of America* 123, pp. 2825–2835.

- Hawkins, S. & J. Midgley (2005). Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association* 35, pp. 183–199.
- Hawkins, S. & N. Nguyen (2004). Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/ in English. *Journal of Phonetics* 32, pp. 199–231.
- Hay, J. B., J. B. Pierrehumbert, A. J. Walker & P. LaShell (2015). Tracking word frequency effects through 130 years of sound change. *Cognition* 139, pp. 83–91.
- Jurasinski, G., F. Koebsch, A. Guenther & S. Beetz (2014). *flux: Flux rate calculation from dynamic closed chamber measurements*. URL <http://CRAN.R-project.org/package=flux>. R package version 0.3-0.
- Kiparsky, P. (1985). Some consequences of Lexical Phonology. *Phonology Yearbook* 2, pp. 83–138.
- Ladd, D. R. (2006). ‘Distinctive phones’ in surface representation. Goldstein, L. M., D. H. Whalen & C. T. Best (eds.), *Laboratory Phonology 8*, Berlin: Mouton de Gruyter, pp. 3–26.
- Ladefoged, P. & I. Maddieson (1996). *The Sounds of the World’s Languages*. Cambridge, MA: Blackwell.
- Lee-Kim, S.-I., L. Davidson & S. Hwang (2013). Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology* 4, pp. 475–511.
- Levelt, W. J., A. Roelofs & A. S. Meyer (1999). A theory of lexical access in speech production. *Behavioral and brain sciences* 22, pp. 1–38.
- Lin, S., P. S. Beddor & A. W. Coetsee (2014). Gestural reduction, lexical frequency, and sound change: A study of post-vocalic /l/. *Laboratory Phonology* 5, pp. 9–36.

- McDougall, K. & F. Nolan (2007). Discrimination of speakers using the formant dynamics of /u:/ in British English. *Proceedings of the 16th International Congress of Phonetic Sciences*, pp. 1825–8.
- Mielke, J. (2013). `tongue_ssanova.r`, URL http://phon.chass.ncsu.edu/manual/tongue_ssanova.r. Functions for SSANOVA comparisons of tongue traces in polar coordinates using `gss`. Accessed March 2015.
- Mielke, J. (2015). An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons). *The Journal of the Acoustical Society of America* 137, pp. 2858–2869.
- Miller, A. L. (2016). Posterior lingual gestures and tongue shape in Mangetti Dune !Xung clicks. *Journal of Phonetics* 55, pp. 119–148.
- Mousikou, P., P. Strycharczuk, A. Turk, K. Rastle & J. Scobbie (2015). Morphological effects on pronunciation. *Proceedings of the 18th International Congress on Phonetic Sciences*, URL <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/proceedings.html>.
- Plag, I., J. Homann & G. Kunter (2015). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* pp. 1–36.
- R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL <http://www.R-project.org>.
- Ramsammy, M. (2015). The life cycle of phonological processes: accounting for dialectal microtypologies. *Language and Linguistics Compass* 9, pp. 33–54.
- Remijsen, B. (2004). Script to measure & check formants, URL http://www.lel.ed.ac.uk/~bert/msr&check_f1f2_indiv_interv.psc. Praat script.
- Rosenfelder, I., J. Fruehwald, K. Evanini & J. Yuan (2011). FAVE (Forced Alignment and Vowel Extraction) Program Suite., URL <http://fave.ling.upenn.edu>.

- Scobbie, J. M. (2007). Interface and overlap in phonetics and phonology. Ramchand, G. & C. Reiss (eds.), *Oxford Handbook of Linguistic Interfaces*, Oxford: Oxford University Press, pp. 17–52.
- Scobbie, J. M. & M. Pouplier (2010). The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction in word-final English /l/. *Journal of Phonetics* 38, pp. 240–259.
- Scobbie, J. M. & J. Stuart-Smith (2008). Quasi-phonemic contrast and the fuzzy inventory: Examples from Scottish English. Avery, P., B. Dresher & K. Rice (eds.), *Contrast in phonology: theory, perception, acquisition*, Berlin: Mouton de Gruyter, pp. 87–113.
- Scobbie, J. M., A. Turk & N. Hewlett (1999). Morphemes, phonetics and lexical items: The case of the Scottish Vowel Length Rule. *Proceedings of the 14th International Congress of Phonetic Sciences.*, International Congress of Phonetic Sciences, vol. 2, pp. 1617–1620.
- Scobbie, J. M., E. Lawson, S. Cowen, J. Cleland & A. A. Wrench (2011). A common co-ordinate system for mid-sagittal articulatory measurement, URL <http://eresearch.qmu.ac.uk/3597/>. QMU CASL Working Papers WP-20.
- Song, J. Y., K. Demuth, K. Evans & S. Shattuck-Hufnagel (2013a). Durational cues to fricative codas in 2-year-olds’ American English: Voicing and morphemic factors. *The Journal of the Acoustical Society of America* 133, pp. 2931–2946.
- Song, J. Y., K. Demuth, S. Shattuck-Hufnagel & L. Ménard (2013b). The effects of coarticulation and morphological complexity on the production of English coda clusters: Acoustic and articulatory evidence from 2-year-olds and adults using ultrasound. *Journal of Phonetics* 41, pp. 281–295.
- Sóskuthy, M., P. Foulkes, B. Haddican, J. Hay & V. Hughes (2015). Word-level distributions and structural factors codetermine GOOSE fronting. *Pro-*

ceedings of the 18th International Congress on Phonetic Sciences, URL <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS1001.pdf>.

Sproat, R. & O. Fujimura (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21, pp. 291–311.

Sugahara, M. & A. Turk (2009). Durational correlates of English sublexical constituent structure. *Phonology* 26, pp. 477–524.

Turk, A. E. & S. Shattuck-Hufnagel (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics* 28, pp. 397–440.

Turton, D. (2014). *Variation in English /l/: Synchronic reflections of the life cycle of phonological processes*. Ph.D. thesis, University of Manchester.

Turton, D. & M. Baranowski (2015). Absence of a blocking r[ɹɫ]?: the presence of /u/-fronting before /l/ in Manchester. Paper presented at 10th UK Language Variation & Change, York.

Uffmann, C. (2012). (Sociolinguistic) variation and phonological representations. Talk presented at The Institute for Linguistics and Language Studies seminar series, University of Manchester.

Wrench, A. & J. M. Scobbie (2016). Queen Margaret University ultrasound, audio and video multichannel recording facility (2008-2016), URL <http://www.qmu.ac.uk/casl/news/WP.htm>. QMU CASL Working Papers WP-24.

Appendix

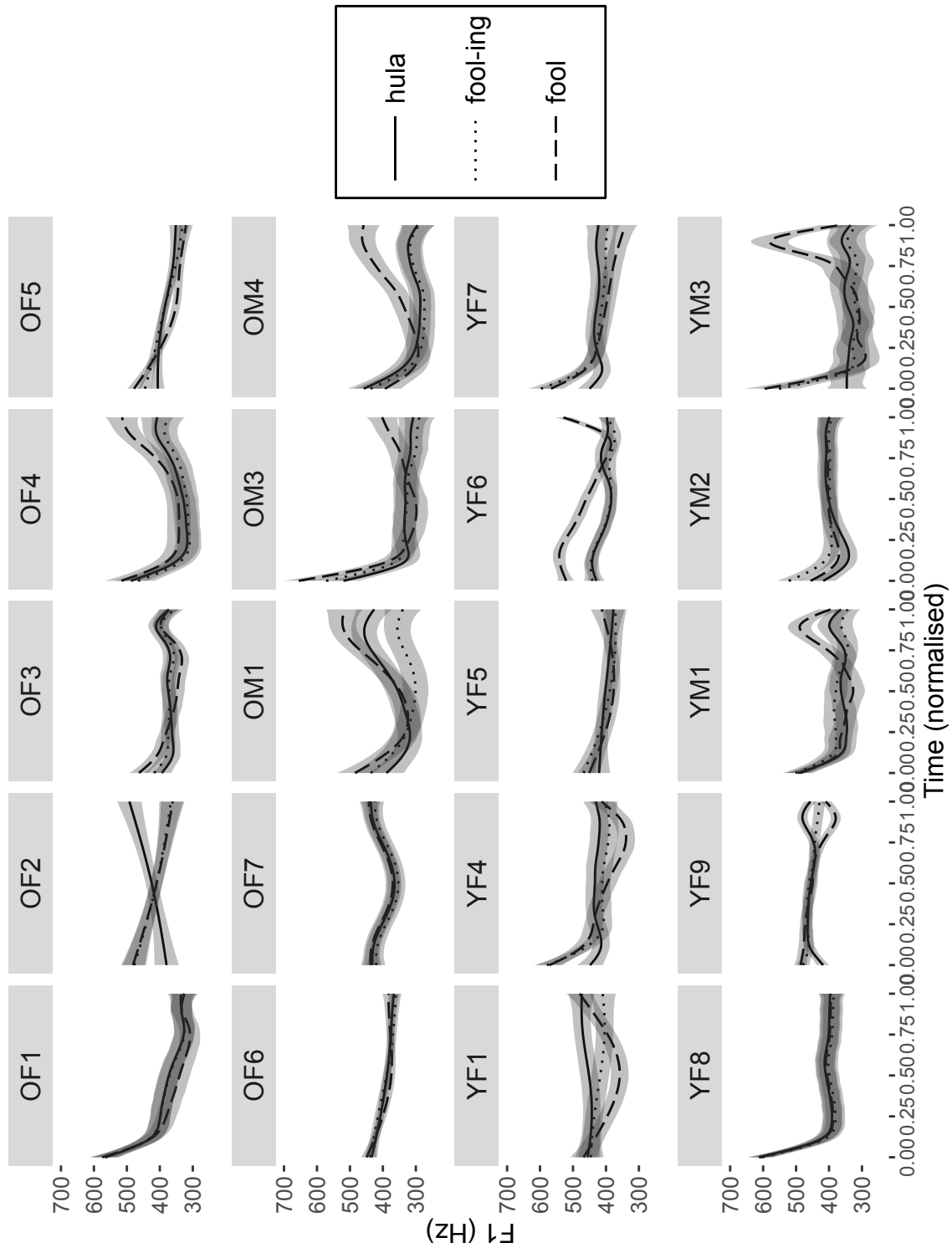


Figure 13: Results of individual speaker SS ANOVA analyses of time-normalised F1, depending on the morpho-syntactic context. Individual speaker codes include information about age group (O=older, Y=younger) and sex (F/M).

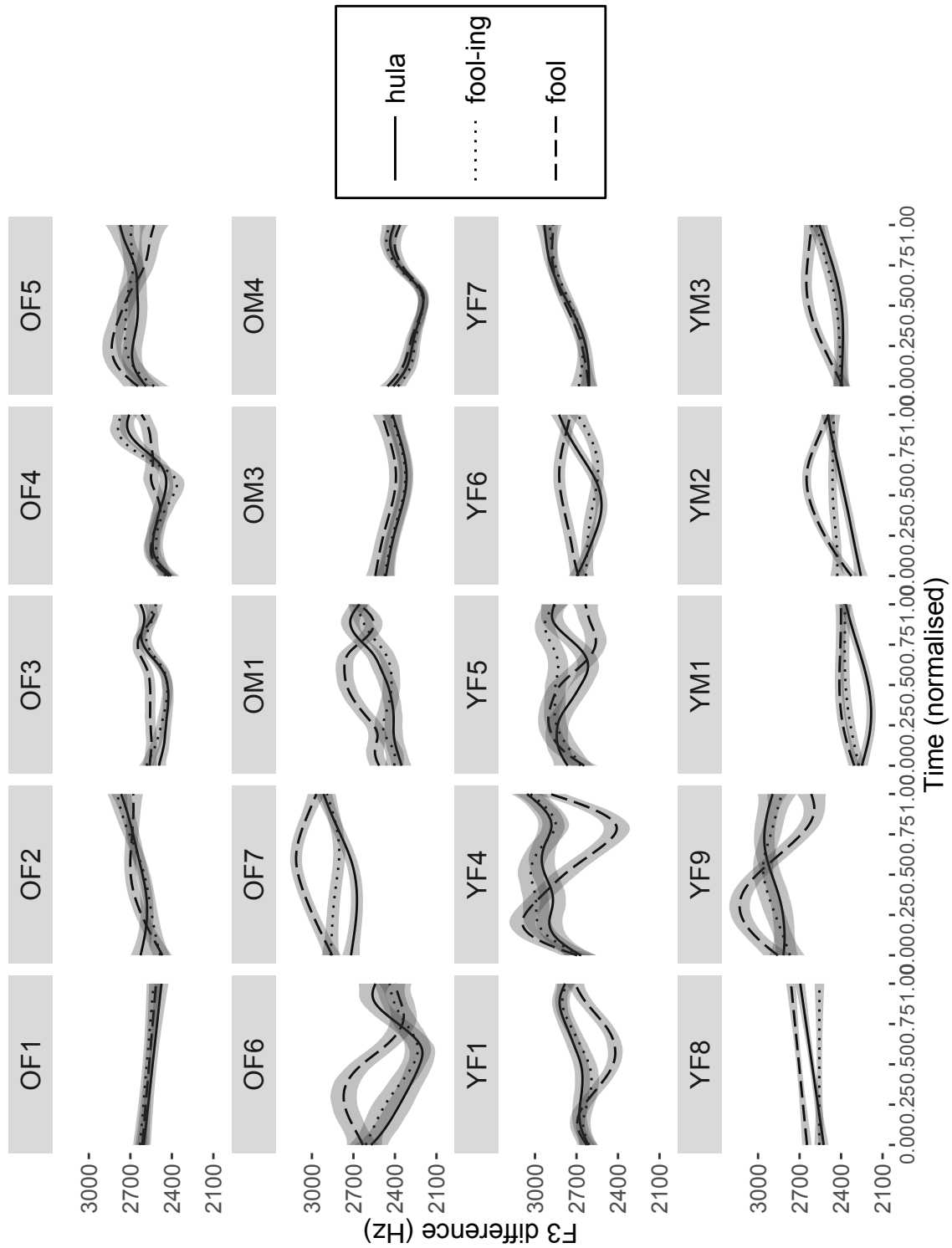


Figure 14: Results of individual speaker SS ANOVA analyses of time-normalised F3, depending on the morpho-syntactic context. Individual speaker codes include information about age group (O=older, Y=younger) and sex (F/M).